DOCUMENT RESUME

ED 190 652 TM 800 449

AUTHOR Stevens, Robert J.: Rosenshine, Barak V.
TITLE Student Ratings of College Instruction: Some Low Inference Variables.
PUB DATE Apr 80
NOTE 20p.: Paper presented at the Annual Meeting of the American Educational Research Association (64th, Boston, MA, April 7-11, 1980).

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Factor Structure: Higher Education: *Rating Scales: *Student Evaluation of Teacher Performance: Teacher Improvement: *Teaching Styles

ABSTRACT

To provide diagnostic feedback for college faculty, a faculty evaluation rating scale was developed and used by college students. The instrument consisted of 2 global items--rate the instructor, rate the course: 5 questions on general teaching characteristic--presentation, enthusiasm, discussion, organization, and personal attention: and 42 low inference items. The latter were specific, operational descriptions of teaching techniques. The five general characteristics correlated .30 or higher with the "rate the instructor" item: they frequently correlated at that level with the "rate the course" item. The low-inference items having the highest correlationi with the general characteristics are listed. The highly significant, partial correlations found between enthusiasm-discussion and presentation-organization suggest that there are not five but only two unique general characteristics: furthermore, both paris of characteristics have some low-inference behaviors in common. Rating forms need consist only of overall evaluation questions, for administrative purposes, and one question for each of two or three general characteristics. From the results of the general characteristics ratings, the instructor receives a list of correlated specific items as feedback. (CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED190652

MAR 1980
RECEIVED
ERIC/TM

Abstract

Student Ratings of College Instruction: Some Low Inference Variables
ROBERT J. STEVENS, University of Illinois
BARAK V. ROSENSHINE, University of Illinois

The objective of this study was to begin to develop a set of specific, low inference items which are related to general characteristic factors commonly used on college instructor rating forms. The results delineate some specific items related to factors of presentation, organization, discussion, enthusiasm and personal attention. These low inference items can be used to provide diagnostic feedback to the instructors concerning their teaching performance, and this feedback could provide the instructors with some specific information on how to improve their performance.

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY
R. Stevens

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

TM 800449

Student Ratings of College Instruction: Some Low Inference Variables

ROBERT J. STEVENS
BARAK V. ROSENSHINE
University of Illinois

Student ratings of instruction are used to evaluate and improve college instruction (Smock and Crooks, 1973). The purpose of this study was to begin to develop a set of specific, low inference items which are correlated with the general characteristic factors commonly used on college instructor rating forms. The low inference items can then be used to provide diagnostic feedback to instructors concerning their teaching performance.

Student ratings have consistently been shown to be a reliable means of evaluation (Costin, Greenough and Menges, 1971; Doyle, 1975; and Feldman, 1978). They have also been shown to be valid. Both overall ratings and particular factor ratings have been positively and significantly correlated with achievement, as measured by content-specific criterion tests (Braskamp, Caulley, and Costin, 1979; Bryson, 1975, Centra, 1977; Frey, 1973, 1976; Gessner, 1973; Marsh, 1977; Marsh, Fleiner and Thomas, 1975; and McKeachie, Lin and Mann, 1971).

However, ratings on factor scores are too vague to provide the instructor with specific, usable feedback. When given feedback on these high inference items, instructors may find it difficult to translate the results into behaviors. Many factor analysis studies have investigated items which are clustered with various factors, but in many cases those items also lack the specificity necessary for providing useful feedback.

Smock and Crooks (1973) categorized items into Level 1, Level 2, and Level 3, as levels of increasing specificity. Level 3 was described as specific items much like Rosenshine's (1970) low inference. It is these low inference items which have the specificity required to give the instructor useful feedback which can be easily translated into behaviors.

In an investigation of these levels of specificity, Brandenburg, Derry and Hengstler (Note 1) used a hierarchical factor analysis on the results from student ratings. They found three levels of specificity, much like those presented by Smock and Crooks.

a) Global - general, summative evaluation.

b) General characteristic - general areas or attributes of instruction.

c) Specific - specific attributes or aspects of teaching.

(Brandenburg, et. al., Note 1, p. 4)

To facilitate readability and ease understanding, these categories will be used throughout the remainder of this paper.

The amount of effect that student ratings can have on teacher effectiveness or teacher behavior has not been fully investigated. There is some evidence that teachers who are presented with accurate and specific feedback can use this information to change their behavior (Good and Brophy, 1973; Moore and Schaut, (Note 3), and Pambookian, 1976). Experimental studies have been done by Centra (Note 2) and McKeachie (1975) using student ratings as the source of specific feedback to college instructors. In both studies the instructors who received specific feedback showed significiant increases in the direction indicated by the feedback, as measured by higher final student ratings,

thus suggesting that specific, usable feedback can help instructors improve their teaching.

Method

In order to develop a set of specific, low inference items which could be used as diagnostic feedback to instructors, the authors developed a rating form containing both global, general characteristics and specific items. The general characteristics chosen were those which were most frequently correlated with achievement in previous research. In attempting to write the new specific items, it was desired that they not only be low inference items, but that they also provide the instructor with some information on which s/he could act. (In writing each item we asked "After reading this item, would the instructor know what to do to change his/her behavior?")

The rating instrument consisted of two global questions (rate the instructor, rate the course), five questions rating general characteristics (presentation, enthusiasm, discussion, organization and personal attention), and 42 specific items. Due to the length of the instrument, it was split into two forms with each form containing the global and general characteristic items, but only half of the specific items, the second half of which were on the other form. The forms used a five point response scheme ("almost always occurred" to "almost never occurred") for the general characteristic items and the specific items. The global items used a five point response scheme of "excellent" to "poor".

The questionnaires were then given to students in botany, educational psychology and music classes at the University of Illinois during

the fall semester 1976 and the spring semester 1977. Each student was randomly assigned to either form A or form B and asked to complete the questionnaire evaluating the instructor of the course. The students were told that their responses would remain anonymous, and that they were not to identify themselves on the answer sheets. The sample sizes for the two forms were 136 and 119 students respectively.

Results

There is some question as to which unit of analysis is most appropriate to use when analyzing student ratings. Although the student is typically the unit of analysis, it has been argued that using the classroom would be more appropriate. "Since the focus of the ratings is the instructor, it might be argued that the classroom is a more appropriate unit of analysis. Accordingly, classroom item means would be the basic data" (Linn, Centra and Tucker, 1975, p. 278). However, a study designed to compare the factors resulting from total group, between group and within group analyses, by Linn, et al. (1975), did not support this argument. Instead the researchers found that the "total group factor solution provided a very good fit to both the between and within covariance matrices" (Linn, et al., p. 288) and thus the "factors from previous total group analyses would be expected to provide good approximations to the between group covariances" (Linn, et al., p. 288). Thus the use of the student as the unit of analysis provides a good approximation of the results of analysis by class, and therefore is used as the unit of analysis in this study.

Initially the results showed a strong correlation between the global ratings and the general characteristic items, suggesting evidence of a

halo effect. The five general characteristics all correlated .30 or better with the "rate the instructor" item on both forms, as well as frequently correlating at that level with the "rate the course" item.

Insert Table 1 Here

In order to eliminate this halo effect upon the students' evaluations of the general characteristic and specific items, a partial correlation was performed, partialling out the two overall evauation items from the correlations of the general characteristic and specific items.

The .01 level of signficance was used as a more stringent criteria for significance to help to eliminate the overlap across the general concepts. This made the results more clear and interpretable, however, some overlap did remain as will be discussed later. The list of the specific items signficiantly correlated with particular general characteristic items (as presented in Table 2) suggest some of the behaviors related to ratings on that general characteristic item.

Insert Table 2 Here

Using these results, and grouping the specific items with those variables with which they had the highest correlation (thus having only one entry per specific item) provides a set of operationalized specific behaviors under each general characteristic (presented below).

Presentation:

The instructor pointed out what was important to learn in each class session.

The instructor summarized the material presented in each class.

The instructor defines students' responsibilities in the course.

The instructor used periodic reviews when making logical transitions.

Personal Attention:

Concern was shown for individual differences.

The instructor provided appropriate material for differing rates of progress.

The instructor checked frequently on students' understanding of the material.

Enthusiasm:

The instructor was a dynamic and energetic person.

The instructor praised the work of the students.

The instructor spoke in a monotone, rarely showing expression in his voice. (negative)

The instructor is clear and concise in presentation and explanation of the material.

The readings were relevant to the course objectives.

Discussion:

The instructor used different methods and materials.

The instructor used student responses/contributions in developing the lesson.

The instructor made an effort to show the interesting nature of the topic.

The instructor used a variety of teaching methods.
The instructor provided alternative ways of learning the course material.

The material was too superficial to adequately develop my skill on concepts. (negative)

The instructor used gestures while teaching

thus the teachers who were rated high on the general characteristic items more frequently exhibited those behaviors listed under that general characteristic. For example, an instructor who was rated high on presentation was one who pointed out what was important to learn, summarized the material presented, defined students' responsibilities and used periodic reviews when making transitions.

The fact that some of these general characteristics also show significant intercorrelations with each other, even after partialling out the overall evaluations, is also of interest. These results, presented in Table 3, suggest that there is not a clear differentiation between these general concepts.

Insert Table 3 Here

The highly significant partial correlations ($p < .01$) foudn between the variables of enthusiasm and discussion, and the variables of presentation and organization, found on both forms provide evidence that there is a great deal of overlap between these general characteristics. Likewise there is evidence that the characteristic of personal attention has some overlap with the other genreal characteristics, in particular the

presentation and enthusiasm variables. (However, these results for personal attention were not consistent across the two forms.) This may then be evidence that there are not five, but rather only two unique factors involved in such ratings. (One study, done by Brandenburg and his associates (Note 1) supports this point of view.) These results suggest factors of presentation/organization, and enthusiasm/discussion, with personal attention not clearly falling in with either one. This result is given further support by the fact that these pairs of variables also have some low inference behaviors in common (See Table 2). These concurrent specific behaviors for the presentation/ organization concepts are:

The instructor spent time in material relevant to course objectives.

The instructor was able to answer questions clearly and concisely.

The instructor pointed out what was important to learn.

The instructor defines students' responsibilities.

The instructor developed eye contact with the students.

Similarly for the enthusiasm/discussion concepts the common items were:

The instructor was a dynamic and energetic person.

The instructor used student responses/contributions in developing the lesson.

And as suggested by the intercorrelations, the personal attention concept had items which were common to enthusiasm:

The instructor was a dynamic and energetic person.

and common to presentation:

The instructor spent time in material relevant to course objectives.

The instructor was adequately prepared for each class.

These results are by no means conclusive on this point, and further study is needed to determine how many factors are involved in these ratings.

Discussion

The results of this study provide a list of specific (low inference) behaviors which are related to students' evaluations on particular characteristics. The specifics can provide the instructor with potentially valuable feedback concerning these student ratings. By providing the instructor with the behaviors related to the general characteristics along with the raw scores (or percentile scores), he will have specific information on which to act. This information can provide some help to the instructor attempting to change his classroom behavior. An instructor who is rated low on a general characteristic item can refer to the specific behavior related to the general characteristic. The instructor can then use this feedback in such a way as to change his behavior in an attempt to improve in the area of that particular general characteristic. For example, if the instructor was rated low on the presentation characteristic, he should attempt to exhibit more frequently the behaviors related to that variable. Therefore the instructor should point out what is important to learn in class; summarize the material presented in class; define students' responsibilities; and so forth. Although this study does not totally exhaust the set of behaviors which may be related to these variables, it does provide an initial step in delineating the low inference behaviors related to these high inference variables.

Another useful result of this study, and studies like it, is that it suggests that rating forms need consist only of overall evaluation questions (for administrative purposes) and one question for each of two or three general characteristic items. From the results of the general characteristic ratings the instructor can be provided with the list of

related specifics as prescriptive feedback. The addition of specific behaviors on the form would be redundant, unless the instructor desired highly specific feedback on those particular behaviors.

Furthermore, this study raises the question of how many distinct factors exist in such rating instruments. Due to the high intercorrelation between pairs of general concept items in this study, it would seem that there are only two or three distinct factors. However, the results are not entirely clear as to how many factors there are, thus warranting further research in this area.

Finally there is a need for experimental studies to utilize the information of this and similar studies. Using the prescriptive feedback, as the experimental condition, similar to those used by Centra (Note 2) and McKeachie (1975), one could assess the true value of these ratings and their results to the teachers being evaluated. In this way it may be determined whether specific, diagnostic feedback to a teacher concerning his teaching performance, as rated by the students, actually influences the teacher in a way to change his behavior.

Table 1

Correlations Between Global Evaluations and General Characteristic Evaluations

Form A	General characteristics:	P	E	D	O	PA
Rate the instructor		.30	.48	.40	.53	.33
Rate the course		.24	.30	.24	.31	.11
Form B						
Rate the instructor		.37	.50	.35	.61	.44
Rate the course		.27	.49	.21	.30	.32

General characteristics: (with item used to measure the concept)

P = Presentation; "The main points of the lecture were clearly understood."

E = Enthusiasm; "The instructor presented the material with enthusiasm."

D = Discussion; "The instructor initiated fruitful and relevant discussions."

O = Organization; "The instructor presented the material in a well-organized fashion."

PA = Personal Attention, "The instructor showed consideration and empathy for the students."

Table 2

Low Inference Items with Significant Partial Correlations with High Inference Variables ($p < .01$)

Presentation	Partial Correlation
The instructor pointed out what is important to learn in each class session.	(.45)
The instructor summarized the material presented in each class.	(.38)
The instructor defines the students' responsibilities in the course.	(.36)
The instructor used periodic reviews when making logical transistions.	(.35)
The instructor developed eye contact with the students.	(.30)
The instructor provided alternative ways of learning the course material.	(.27)
The instructor spent time in material relevant to the course objectives.	(.26)
Concern was shown for individual differences.	(.25)
The instructor was able to answer questions clearly and concisely.	(.25)
The instructor provided practice comprehending course material.	(.25)
Organization:	
The instructor was able to answer questions clearly and concisely.	(.46)
The instructor was adequately prepared for each class.	(.46)
The content was sequenced in logical fashion.	(.45)
The instructor spent time in material relevant to the course objectives.	(.39)
The instructor pointed out what was important to learn in each class session.	(.38)

Table 2 (continued)

The instructor defines the students' responsibilities in the course.	(.31)
The instructor used different methods and materials.	(.29)
The instructor was confident in his presentations.	(.27)
The instructor developed eye contact with the students.	(.27)

particular techniques or styles.

Personal Attention:

Concern was shown for individual differences.	(.43)
The instructor provided appropriate material for differing rates of progress.	(.37)
The instructor developed eye contact with the students.	(.32)
The instructor spent time in material relevant to the course objectives.	(.30)
The instructor provided practice comprehending course material.	(.30)
The instructor was a dynamic and energetic person.	(.29)
The instructor checked frequently on students' understanding of the material.	(.26)
The instructor defines the students' responsibilities in the course.	(.25)

Enthusaism:

The instructor was a dynamic and energetic person.	(.63)
The content was sequenced in logical fashion.	(.32)
The instructor used different methods and materials.	(.31)
The instructor used student responses/contributions in developing the lesson.	(.31)
The instructor was adequately prepared for each class.	(.31)
The instructor praised the work of the students.	(.30)

Table 2 (continued)

The instructor spoke in a monotone, rarely showing expression in his voice.	(-.30)
Concern was shown for individual differences.	(.29)
The instructor was clear and concise in presentation and explanation of the material.	(.29)
The instructor explained the underlying rationale for particular techniques or styles.	(.26)
The readings were relevant for the objectives of the course.	(.26)

Discussion:

The instructor used different methods and materials.	(.49)
The instructor used student responses/contributions in developing the lesson.	(.39)
The instructor was a dynamic and energetic person.	(.37)
The instructor made an effort to show the interesting nature of the topics.	(.37)
The instructor used a variety of teaching methods.	(.33)
The instructor provided alternative ways of learning the course material.	(.30)
The instructor used teacher-made materials.	(.29)
The material was too superficial to adequately develop my skills or concepts.	(-.27)
The instructor used periodic reviews when making logical transitions.	(.27)
The instructor developed eye contact with the students.	(.26)
The instructor used gestures while teaching.	(.26)

Table 3

Partial Correlations Between General Concepts

Form A	P	E	D	O	PA
presentation	1.0				
enthusiasm	.03	1.0			
discussion	.08	.28**	1.0		
organization	.31**	.07	.14	1.0	
personal attention	.19	.11	.07	.10	1.0
Form B					
presentation	1.0				
enthusiasm	.04	1.0			
discussion	.05	.29**	1.0		
organization	.25**	.23**	.07	1.0	
personal attention	.24*	.28**	.04	.24*	1.0

(* $p < .05$; ** $p < .01$)

Reference Notes

1. Brandenburg, D., Derry, S., and Hengstler, D. "Validation of an Item Classification Schema for a Student Rating Item Catalog." Paper presented at NCME annual meeting, March 1978, Toronto, Canada.

2. Centra, J. "Two Studies on the Utility of Student Ratings for Instructional Improvement: 1. The Effectiveness of Student Feedback in Decodifying College Instruction; 2. Self-Ratings of College Teachers: A Comparison with Student Ratings," Princeton NJ: Educational Testing Service, 1972.

3. Moore, J. W. and Schaut, J. "An Evaluation of the Effects of Conceptually Appropriate Feedback on Teacher and Student Behavior," paper presented at the Association for Teacher Education Conference, New Orleans, 1975.

References

Braskamp, L., Caulley, D., and Costin, F., "Student Ratings and Instructor Self-Ratings and Their Relationship to Student Achievement" American Educational Research Journal, 1979, 16:295-306.

Bryson, K. "Teacher Evaluations and Student Learning: A Reexamination" Journal of Educational Research, 1975, 68:12-14.

Centra, J. "Student Ratings of Instruction and Their Relationship to Student Learning" American Educational Research Journal, 1977, 14:17-24.

Costin, F., Greenough, W., and Menges, R. "Student Ratings of College Teaching: Reliability, Validity and Usefulness" Review of Educational Research, 41:511-535.

Doyle, K. Student Evaluation of Instruction, Health & Co.: Lexington, Mass., 1975.

Feldman, K. "Consistency and Variability Among College Students in Rating Their Teachers and Courses: A Review and Analysis" Research in Higher Education, 1978, 9:69-91.

Frey, P. "Student Ratings of Teaching: Validity of Several Rating Factors" Science, 1973, 182:83-85.

Frey, P. "Validity of Student Instructional Ratings" Journal of Higher Education, 1976, 47:327-336.

Gessner, P. "Evaluation of Instruction" Science, 1972, 180:566569.

Good, T., and Brophy, J. "Changing Teacher and Student Behavior: An Empirical Investigation," Journal of Educational Psychology, 1974, 66:390-405.

Linn, R., Centra, J., Tucker, L. "Between, Within, and Total Group Factor Analyses of Student Ratings of Instruction" Multivariate Behavioral Research, 1975, 10:277-288.

Marsh, H. "The Validity of Students' Evaluations: Classroom Evaluations of Instructors Independently Nominated as Best and Worst Teachers by Graduating Seniors," American Educational Research Journal, 1977, 14:441-447.

McKeachie, W., "Changing Teacher Behaviors to Improve Instruction" in Reform, Renewal, Reward D. Allen, M. Melnik and C. Peele (editors) University of Massachusetts: Amherst, Mass., 1975.

McKeachie, W., Lin, Y., and Mann, W. "Student Ratings of Teacher Effectiveness: Validity Studies" American Educational Research Journal, 1971, 8:435-446.

Pambookian, H. "Discrepancy Between Instructor and Student Evaluations of Instruction: Effects on Instruction: Instructional Science, 1976, 5:63-75.

Rosenshine, B. "Evaluation of Classroom Instruction: Review of Educational Research, 1970, 40:279-301.

Smock, H. and Crooks, T. "A Plan for the Comprehensive Evaluation of College Teaching" Journal of Higher Education, 1973, 44:577-586.